Hazard prediction discriminates between novice and experienced drivers

David Crundall

Division of Psychology

School of Social Sciences

Nottingham Trent University, UK

Address for correspondence:

Prof. David Crundall

Division of Psychology

School of Social Sciences

Nottingham Trent University

Burton Street

Nottingham NG1 4BU

e-mail: david.crundall@ntu.ac.uk

Keywords: situational awareness, hazard perception, driving, prediction, What Happens Next?

ABSTRACT

Typical hazard perception tests often confound multiple processes in their responses. The current study tested hazard prediction in isolation to assess whether this component can discriminate between novice and experienced drivers. A variant of the hazard perception test, based on the Situation Awareness Global Assessment Technique, found experienced drivers to outperform novices across three experiments suggesting that the act of predicting an imminent hazard is a crucial part of the hazard-perception process. Furthermore three additional hypotheses were tested in these experiments. First, performance was compared across clips of different length. There was marginal evidence that novice drivers' performance suffered with the longest clips, but experienced drivers' performance did not, suggesting that experienced drivers find hazard prediction less effortful. Secondly, predictive accuracy was found to be dependent on the temporal proximity of visual precursors to the hazard. Thirdly the relationship between the hazard and its precursor was found to be important, with less obvious precursors improving the discrimination between novice and experience drivers. These findings demonstrate that a measure of hazard prediction, which is less confounded by the influence of risk appraisal than simple response time measures, can still discriminate between novice and experienced drivers. Application of this methodology under different conditions can produce insights into the underlying processes that may be at work, whilst also providing an alternative test of driver skill in relation to the detection of hazards.

INTRODUCTION

Hazard perception (HP) has been defined differently by many researchers (cf. Jackson et al., 2009) though one increasingly common description is ‘the ability to predict dangerous situations on the road’ (Wetton, Hill and Horswill, 2013, p65; McKenna and Horswill, 1999; Horswill and McKenna, 2004). This definition fits with attempts by researchers to understand hazard perception within the theoretical framework of situation awareness (Horswill and McKenna, 2004). This theory describes a process of generating and updating a mental model of the current environment as it relates to your goals (e.g. Jeannot et al., 2003; Walker et al., 2009). The most popular version of situation awareness (SA) refers to a linear process that generates a situation model through three stages or levels: (L1) *perception* of elements in the environment, (L2) *comprehension* of their qualities and relevance to current goals, and (L3) *projection* of their locations in space over a suitable timeframe for the task at hand (Endsley, 1988a, 1995; Bolstad et al., 2010). While the precise stages and processes involved might be debated (cf. Vidulich et al., 1994; Gugerty, 2011; Walker et al., 2009), most agree that the prediction of future event states (L3), such as potential hazards on the road, is an important outcome for any instantiation of SA. It is understandable then that this framework has been linked to driver safety and the ability to detect hazards.

However, while many studies make a discursive link between SA and HP, relatively few have used SA as the basis for a study in hazard perception. As Gugerty (2011) pointed out, “research on expert-novice differences for [the main measures of SA are] not available for driving tasks…”p19-6. The current paper addresses this gap, at least in part, by describing three studies which use a variant on the hazard perception test, derived from the most accepted SA measure (the Situation Awareness Global Assessment Technique, or SAGAT).

Before presenting the studies this paper will describe the current approach to hazard perception testing and propose how it can benefit from a test derived from the situation awareness framework. It should be noted however that it is not the intention of this paper to provide support for the Situation Awareness model *per se,* or for the application of SA to the field of hazard perception, but merely to adapt SA techniques to allow better isolation of the prediction component in hazard perception. While these experiments attempt to find evidence of prediction as a differentiator of driver groups, any positive findings could still be viewed as agnostic towards the Situation Awareness model.

The traditional approach to hazard perception

The current UK hazard perception test is based upon decades of research dating back to the 1960s and 70s (e.g. Pelz and Krupat, 1974, Watts and Quimby, 1979). This research was based on the simple underlying hypothesis that safer drivers are more likely to spot hazards earlier than unsafe drivers, and therefore respond to them more quickly. Over the last 40 years this hypothesis has been further unpacked. For instance spotting, comprehending, appraising and responding are all separate aspects of interacting with an on-road hazard, which are further influenced by general driving strategies, caution, and sensation seeking (for a review of these and other factors see Hoswill and McKenna, 2004). Despite this, there has been relatively little effort to explain hazard perception in a broader theoretical framework. One reason for this may be that the majority of research, at least in the previous century, was funded by sponsors who were interested in generating a diagnostic test of driver ability, rather than developing a theoretical basis for HP. It appears however that some researchers have now begun to notice the theoretical lacuna underlying hazard perception. As noted in a factsheet produced by the Netherlands' Institute of Road Safety Research (SWOV, 2010) "Some people are of the opinion that hazard perception is too limited a concept and they prefer to talk about situation awareness", p2. Certainly the 3 levels of *perception*, *comprehension* and *prediction* of future states contained in Endsley's definition of SA (1988a, 1995) appear at first glance to fit well with the different aspects of hazard perception that have been defined by various authors (e.g. Crundall et al., 2008; 2012; Groeger 2000). This would include the ability to first detect potential targets, to then understand their hazardous potential, to link them together in time and space, and then anticipate the most likely hazardous outcome.

To see how these stages might map onto traditional measures of hazard perception one can look at the official UK hazard perception test (a part of the licensing procedure since 2002). Like many HP tests used by researchers, the official UK test requires participants to watch a series of video clips[1] taken from the driver's perspective, and to make a timed response as soon as they perceive a hazard that they would need to avoid by braking or steering. Each official UK clip has a temporal scoring window that falls immediately prior to the full development of each hazard. Responses outside the hazard window (either before or after the window) fail to score anything. However, a response made during the scoring window represents spotting the hazard 'in time to avoid it' and is awarded between 1 and 5 points, with higher points reflecting earlier responses within the window.

The official guidelines for the UK hazard perception test suggest that participants should respond to 'developing hazards' with the following example: "…consider a parked

---

[1] In January 2015 the UK Driver and Vehicle Standards Agency introduced computer-generated clips instead of filmed clips.

vehicle on the side of the road. When you first see it, it is not doing anything; it is just a parked vehicle. If you were to respond to the vehicle at this point, you would not score any marks, but you would not lose any marks. However, when you get closer to the vehicle, you notice that the car's right hand indicator [turn signal] starts to flash. The indicator would lead you to believe that the driver of the vehicle has an intention of moving away, therefore the hazard is now developing and a response at this point would score marks. The indicator coming on is a sign that the parked vehicle has changed its status from a potential hazard into a developing hazard. When you get closer to the vehicle you will probably see the vehicle start to move away from the side of the road; another response should be made at this point"[2] (http://www.nidirect.gov.uk/the-hazard-perception-test-hpt-explained).

The above example hazard has very distinct phases of *potential* (the parked car), *developing* (it indicates) and *actual* hazard states (it moves off). While not explicit in the instructions, it is clear that the official UK hazard perception test is attempting to capture something of the predictive nature of hazard perception, akin to the third level of Situation Awareness.

However it is not clear what the simple response-time measure used in the UK test, and in many tests developed by researchers, is actually measuring. Many factors are likely to influence and confound the response, including individual differences in judging the hazardousness of an event (response criterion), the time required to process the actual hazard, and the level of confirmatory evidence that one requires before making a response (e.g. Deery, 1999).

---

[2] The UK hazard perception test takes the first response within the scoring window as the correct response. Therefore pressing a second time for the hazard should have no effect on one's score, unless the first response was *too fast* and fell before the start of the scoring window.

This multi-faceted nature of hazard processing has been noted by other researchers. For instance, Borowsky and Oron-Gilad (2013) used three separate tests (hazard perception, hazard categorisation and hazardousness ratings) to assess different components of hazard processing, concluding that risk perception related to the likelihood of a collision can affect real-time measures of hazard perception, while risk perception regarding the severity of a possible collision only played a role in hazard processing when used in hindsight (in their ratings and categorisation tests). This suggests that some aspects of risk perception (which relate the hazard to one's own driving skills) can impact on the simple response time measure used in typical hazard perception tests.

A further problem to the traditional hazard perception test, as used in the UK, is that the response is not assessed for accuracy. A participant may press the button for a reason unconnected to the hazard, yet, providing the response falls within the scoring window, they could still receive maximum points. While some researchers have created versions of HP tests that include a measure of accuracy using the mouse pointer or a touch screen to allow localisation of a hazard (Banbury, 2004; Wetton et a., 2010, 2011), these variants may further confound response times by requiring participants to accurately report the location as part of the speeded response (e.g. a mouse click on a location may take longer than a simple button response depending on the size of the target one is trying to click on).

From a diagnostic perspective these problems may not be important. The typical HP test is not concerned with whether differences between safe and unsafe driver groups are due to the prediction of an imminent hazard, the speed of hazard processing, of the level of perceived risk; it simply seeks to differentiate between groups with a gross measure.

For a diagnostic test, this may be all that is required, providing that the test is found to be reliable across time and valid in terms of separating the safe from the unsafe. However research using hazard perception clips remains mixed, with several researchers failing to discriminate between driver groups (Chapman and Underwood, 1998; Sagberg and Bjørnskau, 2006; Borowsky et al., 2010, Underwood, Ngai and Underwood, 2013).

These equivocal results could be due to the varying stimuli used by different research groups (e.g. some use natural hazards while others use staged hazards, despite there being no evidence as to which are better discriminators of driver safety), or perhaps because of the different criteria used to define safe and unsafe drivers, or novice and experienced drivers (Horswill and McKenna, 2004). One further possibility is that the mixed research findings may reflect the varying inputs of the different sub-components to the gross measure of hazard perception (prediction, processing, appraisal etc.). One approach to identify the impact of these different sub-components on hazard perception ability is to isolate each component and test its ability to discriminate between safe and unsafe drivers in isolation. But what would the first possible sub-component be? Perhaps spotting the hazard is the first element in the chain of responding to a particular hazardous event. Certainly other components of the HP response, such as processing, appraisal and confirmation, are likely to come after the initial fixation of the hazard. It has been shown that more experienced drivers are quicker to first fixate hazards than less experienced drivers (e.g. Crundall et al., 2012), so this might seem an obvious starting point. However, the faster eye movements to hazards must be prompted or cued by earlier information. Crundall et al., (2012) referred to these visual cues to upcoming hazards as their *precursors,* and found that experienced drivers were also more likely to fixate these pre-hazard areas of the visual scene. One interpretation of these results is that experienced drivers use these precursors to predict the possibility of a hazard, which allows them to prioritise areas of the visual scene, leading them to fixate developing hazards earlier than less experienced drivers.

On this basis it is understandable why Situation Awareness – where the primary goal is to project near-future situations – is so appealing to hazard perception researchers. The ability to process the hazard once it has been fixated, and to then appraise the risk it poses appropriately are post-SA, or post-prediction, measures (or are, at least, part of a subsequent post-hazard SA iteration).

The question remains whether the act of predicting an immediate on-road event can still discriminate between safe and unsafe drivers once the confounds of post-prediction processes have been removed from the measure. Indeed, it is a possibility that a purer measure of prediction might be a more robust discriminator of hazard perception skill.

The theoretical framework of Situation Awareness has generated a myriad of measurement techniques which could be adapted to the hazard perception field in order to isolate prediction as a behavioural outcome, yet they remain mostly unused in this field (Gugerty, 2011). Not only would the application of SA techniques potentially overcome the problems detailed above, but they would also provide new ways of exploring the theoretical underpinnings of hazard perception.

Can Situation Awareness techniques be applied to hazard perception?

While linking HP to SA seems intuitively appealing, in practice this is hindered by the fact that situation awareness is perhaps as fractionated a concept as hazard perception.

At the start of the century Durso and Gronlund (2000) concluded that the lack of empirical findings in the field of situation awareness fell below the threshold of what was required to develop a coherent theory. Since then, the data may have increased, but consensus has not followed (cf. Salmon and Stanton, 2013). There are a variety of approaches to challenge the predominant *perception, comprehension, prediction* framework proposed by Endsley (1988a; 1995), and researchers debate whether SA is the process of gaining awareness of the situation, or whether it is the product that reflects that awareness (see Salmon et al., 2009, for a concise review). However, as mentioned earlier, this paper is not concerned with the theoretical application of the SA framework to the field of hazard perception *per se*, but merely to investigate the specific role of hazard prediction using tools derived from the SA field. Unfortunately the lack of theoretical consensus in the SA field is mirrored in the variety of techniques that have been used to measure it, with over 30 different methodological approaches identified by Stanton et al., (2005).

The most popular method tends to be a variation on Endsley's (1988b; 1995) Situation Awareness Global Assessment Technique (SAGAT). This method involves the test scenario being stopped (and typically occluded or blanked), followed by probe questions regarding the current and future states of the task environment. This technique has been criticised for being disruptive and relying too heavily on memory (e.g. Durso, Bleckely and Dattel, 2006). These problems are overcome by perhaps the second most common direct assessment of SA; the Situation Present Assessment Method (SPAM; Durso, Truitt, Hackworth, Crutchfield, Nikolic and Moertl, 1995; Durso and Dattel, 2004). This method provides a cue to the participant that a probe question is available. The participant then accepts presentation of the probe question when she feels it is appropriate to do so, and provides an answer while the primary task continues (e.g. the participant chooses when to answer a ringing telephone and then answers the question). The latency to accept a probe question is said to reflect workload (longer latencies imply that the primary task is very demanding at that point in time), while the time taken to answer the question provides an indication of SA (with increasing response times from accepting the probe to providing the answer reflecting poorer SA). Other methods for measuring SA include self-reported or expert-observer ratings (e.g. Taylor, 1990), content analysis of interviews (Walker et al., 2009), verbal commentaries (Salmon, Young and Cornelissen, 2013) and indirect objective measures (or process indices, Gugerty, 2011) including eye movements and physiological measures (e.g. Crundall et al., 2003).

Of all these methodologies, the SAGAT seems most appropriate for investigating the act of prediction in hazard perception. Stopping a video clip immediately prior to the hazard occurring and simply asking ‘what happens next?’ directly targets the predictive component and removes many of the other potential confounds. For instance, the question ‘what happens next?’ does not require the participant to process or judge the subsequent event as having met a particular threshold of hazardousness, which may influence responses to traditional HP tests.

It is important however that a ‘what happens next?’ (WHN) clip is occluded rather than frozen prior to the hazard. Frozen clips do not discriminate between experienced and inexperienced drivers (Vogel et al., 2003), while occluded clips do (Jackson et al., 2009). This suggests that the discriminating factor is the amount and pertinence of information that the driver has when the clip ends. If the clip freezes, the cues to the impending hazard become easily available to everyone, effectively allowing unhurried access to the hazardous precursors followed by a leisurely prediction.

Durso et al., (2006) might argue that the occlusion is disruptive and relies on memory. However the disruption comes at the latest point in the trial, immediately prior to the hazard. These trials are different to the SA tasks that Durso et al. have in mind which are more complex and could last for hours in extreme cases. While a sudden occlusion may disrupt an ongoing SA task such as air traffic control, HP tests are formed of short clips rarely lasting longer than a minute. Occlusion is unlikely to add any additional disruption than is already caused by the transition from one trial to the next in a typical HP test. The criticism of memory is also more likely to hold for long, involved tasks, rather than hazard perception trials. The hazard should be uppermost in participants' memory if they have predicted its impending appearance. One final concern might be that the sudden occlusion prompts participants to lower their threshold for predicting a hazard, effectively up-grading a possible hazard precursor into a definite hazard precursor, however the inclusion of 'non-hazard' trials should hopefully mitigate this.

Identifying prediction as a key element of hazard perception

The main aim of this study is to extend the limited literature that identifies prediction as an important component of hazard perception, by identifying whether the ability to predict an imminent hazard successfully discriminates between novice and experienced drivers. If experienced drivers consistently perform better than novices across a number of experiments, then one might conclude that novices' deficiencies in the measured skill may relate to their increased crash risk compared to more experienced drivers, as typically revealed in national crash statistics (e.g. McKenna and Crick, 1994; Crundall and Underwood, 1998; Horswill and McKenna, 2004). Jackson, Chapman and Crundall (2009) found experienced drivers to be more accurate than novices at predicting hazardous events in occluded clips. Participants had to answer three questions: 'What was the source of the hazard?', 'Where is the source of the hazard?' and 'What happens next?'. Thus a participant could answer 'The pedestrian… on the left pavement… is about to cross between two parked cars into my path' in order to score full points providing each answer was unambiguous (2 points per question). Other studies have demonstrated that such prediction tasks can be useful in training hazard perception skills (McKenna and Crick, 1997), though they might not be as potent as other interventions such as commentary training (Wetton et al., 2013). Recently, the 'What Happens Next?'(WHN) methodology of Jackson et al. (2009) has been refined for the development of a Spanish hazard perception test (Castro et al., 2014), and has also been found to discriminate between driver groups in Malaysia (Lim, Sheppard and Crundall, 2014). Despite the promising start for prediction studies of hazard perception, the literature remains scant. The primary purpose of this paper is to add to the evidence base by assessing whether the ability to predict an upcoming hazard can discriminate between driver groups repeatedly across a series of experiments, demonstrating the importance of pre-hazard processing when isolated from the confounding influences of post-prediction processes such as appraisal.

Furthermore, the use of a hazard prediction measure allows a series of additional theoretical questions to be investigated. Specifically this paper is concerned with the effortful nature of sustained hazard perception (do longer clips require more sustained attention, and does this differentially affect novice and experienced drivers?), the impact of occluding a clip at different temporal distances from the hazard (will a greater separation between occlusion and hazard better discriminate between novice and experienced drivers?), and whether the nature of the hazardous event influences predictive accuracy (separating hazards into Behavioural and Environmental Prediction hazards; Crundall et al., 2012). Each of these questions will be expanded in subsequent sections, though the over-arching hypothesis is that experienced drivers will be better able to predict the subsequent hazard than novice drivers.

Experiment 1: The effortful nature of prediction

Within the SA framework, working memory is required for encoding information, prioritizing potential hazard sources for continuous monitoring, and creating and updating predictions based on numerous elements in the scene, all of which would argue for an effortful process (e.g. Johannsdottir and Herdman, 2010). Gutzwiller and Clegg (2013) found the strongest positive relationship between working memory capacity and SA at level 3: *prediction*. This fits with the notion that hazard perception via prediction requires attentional resources. However there is also evidence that the link between working memory performance and SA diminishes with increasing task experience (e.g. Sohn & Doane, 2004), with experts relying more on long term working memory representations of context, such as chunks (Chase and Simon, 1973), retrieval structures (Ericsson and Kintsch, 1995) or templates (Gobet and Simon, 1996). Chunking, even in working memory, seems not to require executive control (Baddeley, Hitch and Allen, 2009), so one might therefore predict that experienced drivers, with a host of hazardous templates stored in long term memory, should find it less effortful to predict upcoming hazards.

Despite this, the evidence suggests that even highly experienced drivers' abilities to spot hazards can be impacted by a secondary task (e.g. Kim et al., 2013; Divekar et al., 2012). For instance, Savage et al., (2013) found significant slowing of hazard response times, reduced spread of search and increased frontal lobe activity in participants (aged 18-34) watching HP clips while engaging with a secondary task. Only one study however (to this author's knowledge) has assessed secondary task impact on typical HP test responses across novice and experienced driver groups. McKenna and Farrand (1999) demonstrated that random letter generation had a greater detrimental effect upon experienced drivers' response times than novices', effectively reducing performance in both groups to the same level. Thus it appears that, while theoretically one might predict that performance on a HP test should require less effort with greater driving experience, the evidence argues that HP requires equal if not greater efforts from more experienced drivers.

This suggests that either experienced drivers do not use stored hazard templates to help predict upcoming events, or that it was the post-predictive elements of the responses in McKenna and Farrand's study (processing, appraisal, etc.) which were susceptible to secondary task disruption.

A further complication in understanding how effortful HP might be is the use of secondary tasks which might encourage systematically different strategies in different driver groups. A more naturalistic alternative might be to vary the length of hazard clips to identify whether there is a time-on-task decrement. By varying the length of hazard perception clips (while maintaining the upper limit at one minute to prevent monotony and fatigue), a degradation in performance with longer clips would suggest a depletion of attentional resources (cf. Warm et al., 2008). This time-on-task approach provides a more naturalistic measure of the effort required for hazard perception compared to the introduction of an artificial secondary task. This approach can also inform the design of future HP tests, as little empirical consideration has been given to the length of hazard perception clips.

It should also be noted that the longer a trial runs for without the appearance of a hazard, the greater the likelihood that the hazard is about to appear (typical HP tests contain few, if any, target-free trials). One might therefore imagine that participants would invest more heavily in trying to predict the hazard in the latter stages of long clips. Any time-on-task decrement in performance is thus more likely to stem from depleted resources rather than motivation to complete the task.

On these bases one might predict that an effortful process should demonstrate a decline in prediction accuracy with longer clips, while a non-effortful (or minimally-effortful) process should not show such a decrement, or may even show an improvement if participants concentrate harder towards the end of long clips as suggested above. If the findings of McKenna and Farrand (1999) apply to the prediction element of hazard perception, rather than the post-prediction processes also contained in their response time measure, the decrement should be noted for both experienced and novice drivers, and possibly be greater for the experienced drivers. If however McKenna and Farrand's results are primarily derived from post-prediction effort, then experienced drivers should suffer less degradation. Finally, underlying this experiment is the over-arching hypothesis that prediction accuracy should discriminate between novice and experienced drivers.

In order to test these hypotheses a series of occluded hazard prediction clips (with the video image immediately disappearing prior to the hazard) of different lengths (short, intermediate and long), were randomly presented to novice and experienced drivers who were asked to predict what happens next following occlusion.

Method

Participants: Thirty drivers were recruited all of whom had a full UK driving license (23 female, mean age = 20.8 years). Fifteen were classed as novice drivers who had been driving for less than three years (mean driving experience since test = 2.1 years). The remaining 15 were classed as experienced drivers, having held their full driving license for a minimum of 3 years (mean driving experience since test = 3.7 years). All participants were undergraduate students with normal or corrected to normal vision.

Apparatus and Stimuli: Thirty hazardous and 10 non-hazardous clips were drawn from a database of clips, some of which have been used in previous studies. All clips were filmed from either a wide-screen Sony or Panasonic high definition digital video camera, from a suction mount fixed to the inside of a windscreen in a moving vehicle. The clips included both staged and naturally occurring hazards. A wide variety of road ways were depicted in the clips including urban, suburban, dual carriageway and motorway roads, across a mix of weather conditions (sunny, overcast, light mist). Each of the hazardous clips was edited to cut to a black screen immediately prior to the hazard occurring. A typical clip would contain hazardous precursors (clues to the impending hazard) that would allow participants to correctly predict what would happen next if they were looking in the correct location on the screen prior to occlusion. For instance, in Figure 1, the presence of parked vehicles and a pedestrian (zebra) crossing might make one prioritize this area of the scene. The pedestrian may then be noticed and monitored until it is obvious that she is going to become the hazard.

Clips were edited to be either *long* (with a mean length of 44s), *intermediate* (with a mean of 24s), or *short* (with a mean of 10s). Clip length varied within the classifications to prevent participants from detecting the manipulation. Ten non-hazard clips of variable length were also included in the study. Non-hazard clips were randomly selected from the database and were not chosen on the basis of clip length.

Clips were displayed on a 24 inch iMac computer with bespoke software which randomized the clips and presented a black screen as soon as each clip ended. A chin rest ensured a viewing distance of 60 cm.

Design: A 2 x 3 mixed design was employed with the between-group factor of driving experience (novice vs. experienced drivers) and the within-group factor of clip length (long, intermediate and short).

The main dependent variable was the participants' accuracy at identifying the upcoming hazard on the basis of their responses to three questions following each video clip. The three questions (taken from Jackson et al., 2009) asked participants to report the source of the hazard (e.g. "the pedestrian"), the location of the hazard source (e.g. "walking along the pavement on the left"), and then to predict what would have happened next if the screen had not turned black (e.g. "she was about to step into the road without looking"). Clearly correct answers to these three questions were awarded 2 points each. Clearly incorrect answers were awarded zero. On rare occasions where the answer was ambiguous, participants might receive a single point for a question. Thus participants could score a maximum of 6 points across 3 questions for correctly identify and predicting each hazard. If participants incorrectly said that no hazard was about to occur then they scored zero for that trial.

The presentation of 30 hazard clips and 10 non-hazard clips was randomized for all participants within a single block. Non-hazard clips were included to reduce participant guessing.

Procedure: Participants initially filled in a demographic questionnaire which included details of their driving experience and accident history. Participants were then seated 60 cm from the screen and instructed in the task. They were told that they would be presented with videos taken from the driver's perspective and that they should view them as if they were the driver. They were instructed that each clip would end abruptly with a black screen, and then they would be asked three questions pertaining to the any potential hazard source (“What was the source of the hazard”), its location (“Where was the hazard located?”) and “what happens next?”. Participants provided verbal responses to these questions. A hazard was defined as an object or event present in the road environment that could increase the potential risk of a crash occurring without a suitable evasive manoeuvre such as braking or steering to avoid a collision. Participants viewed two practice clips, and gave answers to all questions. If the participants did not perform as expected, the experimenter explained the requirements further.

Results

The summed scores for all questions were turned into percentages for each participant and were analysed in a 2x3 mixed ANOVA. There was a main effect of driver group (F(1,28) = 45.4, MSe = 92.7, $p<0.001$) with experienced drivers performing better than novices (77.8% vs. 54.1%). A main effect of clip length was also found (F(2,56) = 4.4, MSe = 73.7, $p<0.05$). Repeated contrasts revealed that the long clips were responded to less correctly than the intermediate length clips (62.4% vs. 68.8.2%; F(1,28) = 7.4, MSe = 167.6, $p < 0.05$), though there was no difference between the intermediate and short length clips (68.8% vs. 66.6%). The omnibus interaction did not reach conventional levels of statistical significance (F(2,56) = 2.8, MSe = 73.7, $p = 0.07$), though the planned repeated contrasts identified that the novices’ performance was degraded more than the experienced drivers’ performance in the long clips compared to the intermediate clips (F(1,28) = 4.3, MSe = 167.6, $p < 0.05$; see Figure 2).

Discussion

The findings of this study support the methodology proposed by Jackson et al. (2009) to use occluded hazard perception clips to discriminate between experienced and novice drivers. Similar to the results of Jackson et al., there is a significant difference between the novice and experienced drivers' performance on this task. Calculating Cohen's d for the difference between the two groups gives a value of 2.5. Cohen (1988) suggested that a value of just 0.8 can be considered a large effect.

The effect of the clip length on the two groups' performance was apparent in the marginal omnibus interaction, and more so in the planned contrasts (which are not reliant on a significant omnibus effect). The length of the clip has no appreciable effect on the experienced drivers, but for the novice drivers there is an apparent degradation in performance in the longest clips.

This result does not fit with the findings of McKenna and Farrand (1999) who argued that hazard perception is perhaps even more effortful for experienced drivers than novice drivers, as evidenced by the greater degradation they suffer during a secondary task. This inconsistency may reflect the possibility that it is the post-prediction processes that require the most effort (e.g. judging the relative hazardousness of the event). These post-predictive processes would have confounded the simple response-time measures used by McKenna and Farrand. When the predictive component of hazard perception was isolated in the current study, it was found to be more impervious to time-on-task decrements in the experienced drivers, in keeping with the suggestion that experienced drivers might use long-term memory templates to reduce the demand of hazard prediction.

Alternatively, one might question the choice of random letter generation as McKenna and Farrrand's secondary task. For instance there is recent evidence that the generation of random responses is linked to the direction of a participant’s gaze (e.g. Grade, Lefèvre and Pessenti, 2013). If eye movements were influenced by random letter generation in McKenna and Farrand’s study, this could have had a disproportionate effect upon the experienced drivers who have been noted to have better eye movement strategies than novices (e.g. Chapman and Underwood, 1998). Without a complete understanding of the ways in which a secondary task might influence primary task performance, it is arguable that time-on-task decrements provide a more naturalistic alternative.

Regardless of the source of the discrepancy between the two studies, the current results suggest that experienced drivers are both better than novices overall, and less susceptible to a time-on-task decrement on the ability to predict a hazard.

Experiment 2: Manipulating the point of occlusion

The amount of information contained within a *hazard precursor* (the most obvious clue to the appearance of the subsequent hazard) will change over time, increasing as the hazardous event approaches. When a precursor first appears (e.g. a pedestrian pushing a bicycle along a pavement) it may contain relatively little information about the upcoming hazard, and may therefore be given a lower priority by a driver than other more pressing potential precursors. As the clip continues however, the driver may notice a zebra crossing (a marked crosswalk) ahead. She may also note how the pedestrian with the bicycle is looking over her shoulder, and how the trajectory of the pedestrian changes, as if heading towards the zebra crossing. Thus the precursor shifts along an information continuum, starting with little evidence for the upcoming hazard, before accruing some subtle predictive cues (e.g. the pedestrian’s quick glance over the shoulder), which increase in relevance and salience until the cues are relatively obvious (e.g. a pedestrian trajectory that is unmistakably heading for the crossing).

It has previously been noted that learner drivers are less likely to look at hazardous precursors than more experienced drivers, and if they do look at such precursors they are typically slower to do so in the majority of cases than experienced drivers (Crundall et al., 2012). Thus experienced drivers have earlier access to some precursor information, though this information is going to be of less value than later precursor information (which is temporally closer to the actual hazard). Are the experienced drivers actively processing the hazard precursors with these early fixations (i.e. making predictions) or are they merely fixating these precursors as part of a general search strategy that has developed with experience (e.g. Underwood et al., 2002)? In the latter case, an early fixation on a precursor may find little informative value to recommend priority inclusion in the current prediction process.

Certainly it is intuitively appealing to interpret the early fixations of experienced drivers on some precursors (Crundall et al., 2012) as reflecting early engagement with predicting the upcoming hazard. However it is also possible that experienced drivers take relatively late information from precursors but still process it faster than novices (e.g. Chapman and Underwood, 1998). The current experiment was therefore undertaken to test whether experienced drivers can actually make use of early subtle cues from precursors.

A selection of hazard prediction clips were edited with three different occlusion points: *early*, *intermediate*, and *late*, with the late occlusion points being temporally closest to the hazard (they were the default occlusion points used in experiment 1). If experienced drivers can extract information from earlier, more subtle precursor cues, then one might predict that the discrimination between novice and experienced drivers should be maintained, or even improved with earlier occlusion points (as novices have very little experience of early subtle cues, due to late fixation of precursors). Alternatively, if the information contained in early precursors is too impoverished, one might predict that novice and experienced driver performance on the prediction test should converge at a nadir.

Method

Participants: Forty-two drivers were recruited, all of whom had a full or provisional UK driving license. Twenty-one novice drivers had held a full driving license for less than three years (16 female, mean age = 20.5 years, mean driving experience since test = 1.8 years). The remaining 21 participants were classed as experienced (14 female, mean age = 23.9 years, mean driving experience since test = 6.4 years). All participants were university students with normal or corrected to normal vision.

Apparatus and Stimuli: Thirty hazardous and 10 non-hazardous clips between 15 and 30 seconds long were chosen for inclusion. Three versions of each hazard clip were created with early, intermediate and late end points. Early end points occurred when the precursor to the hazard is first visible (approximately 1250 ms before the late end point). An intermediate end point occurred while the precursor was progressing towards becoming a hazard (approximately 800 ms on average before the late end point). A late end point occurs immediately prior to the hazard, and was the default choice used in Experiment 1. Figure 1 shows the three different end points for a particular hazard. In this clip the early end point occurs when the head of the pedestrian (the immediate precursor to the hazard) is first visible. The intermediate end point occurs as the pedestrian moves towards the road edge, and the late end point occurs immediately prior to the pedestrian stepping into the road.

The apparatus was the same as that used in Experiment 1.

Design: A 2 x 3 mixed design was employed with the between-group factor of driving experience (novice vs. experienced drivers) and the within-group factor of clip end point (early, intermediate and late, with late end points being temporally closest to the hazard). Groups of clips with different end points were rotated across participants, such that participant 1 was presented with clips 1-10 with an early end point, clips 11-20 with an intermediate end point, and clips 21-30 with a late end point (presented randomly within a single block together with 10 non-hazard clips). Participant 2 was then presented with clips 1-10 with an intermediate end point, clips 11-20 with a late end point, and clips 21-30 with an early end point, and so on.

The main dependent variable was the participants' accuracy at identifying the upcoming hazard on the basis of their responses to three questions following each video clip ("What is the source of the hazard?', 'Where is the source of the hazard located?', 'What happens next?'). In a modification to the previous design, participants were also asked to rate their confidence in their answer on a five-point scale (with higher scores reflecting greater confidence). As in Experiment 1, participants could score a maximum of six points for a correct prediction, down to zero points for incorrect predictions (including incorrectly reporting that 'no hazard' was about to occur).

The presentation of 30 hazard clips and 10 non-hazard clips was randomized for all participants. All participants saw the same non-hazard clips which did not vary by end point and were included in order to reduce participant guessing.

The procedure was identical to that used in Experiment 1, with the exception of the inclusion of a confidence scale following each trial.

Results

Accuracy of participants' predictions were calculated by summing the scores of each clip (out a maximum of 6) for each end point, which were then converted into percentages for a 2 x 3 mixed ANOVA (Figure 3, top panel). Experienced drivers were more accurate than novice drivers (79.9% vs. 69.5%; $F(1,40) = 9.9$, $MSe = 115$, $p < 0.005$) and the main effect of clip end point was also significant ($F(2,80) = 76.3$, $MSe = 142$, $p < 0.001$). Repeated contrasts showed that clips with late end points were responded to more accurately than intermediate end points (88.1% vs. 79.1%; $F(1,40) = 22.2$, $MSe = 152$, $p < 0.001$), while intermediate clips were more accurately responded to than clips with early end points (79.1% vs. 56.9%; $F(1,40) = 66.5$, $MSe = 312$, $p < 0.001$). The interaction between driver group and end point was not significant.

A second 2 x 3 mixed ANOVA was conducted on the mean confidence ratings of participants as to whether their predicted hazard was indeed going to occur (taken from a 1-5 scale where 5 is very confident). A main effect of clip end point was found ($F (2, 80) = 33.2$, $MSe = 0.347$, $p < 0.001$; Figure 3, bottom panel). Repeated contrasts revealed that clips with late end points (which finished immediately prior to hazard onset) produced higher confidence ratings than clips with intermediate end points (4.3 vs. 4.0, respectively; $F(1,40) = 15.8$, $MSe = 0.1$, $p < 0.001$), which in turn received higher confidence ratings than clips with early end points (4.0 vs. 3.7; $F(1,40) = 24.9$, $MSe = 0.2$, $p < 0.001$). Neither the group effect nor the interaction was significant. Confidence ratings to the non-hazardous clips were identical across groups, and similar to the ratings given to the earliest end-point clips (3.7).

Discussion

The findings of Experiment 2 support the group effect noted in Experiment 1, with experienced drivers out-performing the novice drivers in successfully predicting the hazard. The size of this effect is much less than that noted in the first study (Cohen's d = 0.97), though is still considered to be a large effect according the rules of thumb proposed by Cohen (1988). It was also found that occluding the clip earlier in time (further away from the onset of the hazard) reduced prediction accuracy in both groups. While this demonstrates that information immediately prior to the occluded hazard (in clips with a late end point) is vital for evoking high levels of accuracy, it also suggests that information which appears an average of 1200 ms before the hazard onset (the mean end point of the early-termination clips) is still sufficient to produce an experiential difference. The lack of interaction between group and end point suggests that experienced drivers do not benefit disproportionately more than novice drivers from this early hazard precursor information, but neither do they suffer more.

Confidence ratings follow a similar pattern to accuracy across the three clip end points, with earlier end points receiving lower confidence ratings than later end points. The increased accuracy of experienced drivers over novice drivers is not reflected in the confidence ratings however. If confidence ratings were turned into a percentage of the confidence scale, it would suggest that all drivers are over-confident when compared to their actual prediction accuracy, however novice drivers appear the most over-confident across all endpoint conditions, with the greatest discrepancy between confidence and performance appearing in the early end-point clips.

The results support the suggestion above that early fixation of hazard precursors may reflect early engagement with the hazard prediction process. Certainly there appears to be sufficient information within the precursor to result in the superior performance of the experienced drivers. This of course does not imply that an experienced driver forms an accurate prediction during the earliest stages of a precursor when under natural conditions. The sudden occlusion forces drivers to extrapolate from the clues that they have. The confidence ratings clearly demonstrate that experienced drivers do not feel comfortable predicting a hazard on such subtle cues. Nonetheless they must have acquired some information on which to base their decisions and to maintain their performance above that of the novices.

Experiment 3: investigating the relationship between precursor and hazard

Relatively few HP studies have concerned themselves with the relationship between the hazard precursor and the hazard itself. Of those few studies the results are mixed. For instance, Pradhan et al., (2005) found that novice drivers were less likely to look at potential sources of hazard (i.e. potential precursors) than more experienced drivers. Similarly Borowsky, Shinar and Oron-Gilad (2010) found that eye movements to potential hazard sources were more discriminative than fixations on actual hazards between driver groups. In a follow on study, Borowsky and Oron-Gilad (2013) compared ‘materialised’ and ‘hidden unmaterialised’ hazards and reported that their novice drivers had particular problems spotting hazards when they were hidden by the environment. However, another recent study failed to find any interaction between hazard precursor and driver group, when comparing abrupt hazards with gradual-onset hazards across novice and experienced car drivers (Underwood, Ngai, and Underwood, 2013).

In a recent simulator study (Crundall et al., 2012) a number of different relationships between precursors and hazards were posited. The two main ones were termed Behavioural Prediction hazards (BP) and Environmental Prediction hazards (EP). BP hazards are characterized by a continuity between the precursor and the hazard. For instance the example of the pedestrian pushing a bicycle towards a zebra crossing requires the driver to predict the hazard on the basis of the *behaviour* of the pedestrian. If the pedestrian glances over her shoulder, this behaviour may evoke an accurate prediction that she will cross the road (becoming the hazard). However, if we envision a scenario where a pedestrian emerges from behind a high-sided vehicle, then the pedestrian is immediately considered to be a hazard as soon as the driver can see her. What possible precursor could flag the appearance of such a sudden hazard? Crundall et al. (2012) proposed that highly experienced drivers should be aware that a high-sided parked vehicle could mask a potential pedestrian and therefore will devote attention to this area of the *environment*. The experienced driver better understands the role of the environment (akin to having a deeper knowledge of the contextual structure than the novices, cf. Chi et al., 1981), and thus uses the environment to predict the potential hazard. This distinction is loosely similar to the gradual and abrupt distinction used by Underwood et al. (2013), the materialised and hidden-unmaterialised hazards of Borowsky and Oron-Gilad (2013), and the overt and covert latent hazards used by Vlakveld et al. (2011).

The results of Crundall et al. (2012) demonstrated that learner drivers were slower to fixate all types of hazards and most types of precursors. However all driver groups were particularly slow, and less likely, to fixate the EP precursors. This ostensibly suggests that the relationship between the BP precursor and hazard is stronger than that of the EP precursor and hazard. Furthermore one might argue that the EP precursor is of little use to drivers, as even driving instructors fixated these precursors later than other precursors (Crundall et al., 2012). However, while the relationship between an EP precursor and its hazard might be less direct than with a BP hazard, the EP precursor is more temporally specific. For instance, a pedestrian visible on the pavement for some time may need to be fixated on a regular basis as the driver is unsure when she will step into the road. However with an EP precursor (such as a parked van) there is no need to monitor it continuously on approach, as an emergent pedestrian would only be a hazard if the driver was close enough to the van. Thus the parked van need only be fixated when the driver gets so close that they would need to take evasive action should an EP hazard occur. Thus is it possible that the late fixations of experienced drivers and driving instructors on EP precursors mask the fact that they are aware that the van is there (perhaps monitored through their superior useful field of view; Crundall et al., 1999, 2002), and only choose to fixate it when its proximity makes it relevant.

While it is intuitively appealing to suggest that the indirect link between EP precursors and hazards should result in these hazards being harder to detect, the eye movement analyses of Crundall et al. (2012) are equivocal in this regard. In order to assess the relative difficulty of predicting BP and EP hazards a third experiment was undertaken with hazard prediction clips categorized according to the nature of their predominant precursor. It was predicted that, first, the group effect noted in the two previous studies would prevail, supporting the role of prediction as a key discriminator of experienced and novice drivers. Secondly, it was predicted that EP hazards should be more difficult to predict than BP hazards, due to the indirect link between precursor and hazard, and finally, that novice drivers should suffer more than experienced drivers from this indirect link.

Method

Participants: Thirty drivers were recruited all of whom had a full or provisional UK driving license. Fifteen were classed as novice drivers who either had a provisional license or had only held a full license for less than one year (13 female, mean driving experience since test = 0.9 years, mean age = 20.7 years). The remaining 15 were classed as experienced drivers, having held their full driving license for a minimum of 3 years (10 female, mean driving experience since test = 3.4 years, mean age = 20.8 years). All participants were undergraduate students, with normal or corrected to normal vision.

Apparatus and Stimuli: The apparatus was identical to that used in experiment 1. The 30 video clips were selected from the database of clips. Twenty clips contained a hazard (10 leading to Behavioural Prediction (BP) hazards and 10 leading to Environmental Prediction (EP) hazards), and were edited for the current study such that they ended immediately prior to the onset of the hazard, but with enough information so that a participant who is looking in the appropriate place at the appropriate time should be able to identify the upcoming hazard (see Figure 4). BP hazards included a girl being pushed into the road by friends, a pedestrian stepping into the road without looking, a vehicle cutting into your lane, and oncoming traffic turning across your path into a side road. EP hazards included a pedestrian stepping out from between parked cars, a car reversing from a driveway obscured by parked vehicles, oncoming traffic hidden by a maintenance truck, and a pedestrian stepping out from behind a parked bus. The ten non-hazard clips were edited to points where precursors might lead drivers to expect a hazard (e.g. a car waiting to pull out from a side road, but it never does). For all 30 clips, film length varied between 5 and 56 seconds.

Design: A 2 x 2 mixed design was employed with the between-group factor of driving experience (novice vs. experienced drivers) and the within-group factor of hazard type, which included Behavioural Prediction hazards (BP) and Environmental Prediction hazards (EP). BP hazards are typified by having the same precursor as the actual hazard, with hazard onset being defined in terms of the temporal and spatial locations of the target object (e.g. a pedestrian walking on the pavement may be the behavioural precursor, but when the pedestrian steps into the road she becomes the hazard). EP hazards have an indirect link between precursor and hazard, often having the hazard obscured by part of the environment (e.g. a parked van might act as the environmental precursor for an obscured pedestrian who steps into the road). The primary dependent measure was participant accuracy in predicting the subsequent hazard, using the scoring scale from experiment 1 (with a maximum score of 6 for each clip, across 3 questions).

As in experiment 2, participants were also required to rate their confidence in their prediction on a 1-5 scale, with a score of 5 demonstrating high confidence that the hazard will occur.

All video clips were presented in a random order, and in addition to the 20 experimental clips (10 BP and 10 EP clips), ten non-hazard clips were also included.

The procedure was identical to that used in Experiment 1 beyond the inclusion of the confidence ratings.

Results

Accuracy scores for each clip (out of a maximum of 6) were summed for the 10 BP and 10 EP clips, and then converted into participant percentages for a 2 x 2 mixed ANOVA. BP hazards were responded to more accurately than EP hazards (79.7% vs. 61.4%; F(1,28) = 24.6, MSe = 204, $p < 0.001$) and experienced drivers were more accurate than novices (81.6% vs. 59.6%; F(1,28) = 17.4, MSe = 419, $p < 0.001$). An interaction between these two factors was also noted (F(1,28) = 11.3, MSe = 204, $p < 0.005$). As can be seen in Figure 5, experienced drivers appear equally good at predicting EP and BP hazards [t(14) = 1.1, $p > 0.1$], but novice drivers are much worse at predicting EP hazards compared to BP hazards (t(14) = 6.2, $p < 0.001$). Furthermore, there is only marginal evidence that experienced drivers are better than novice drivers when predicting BP hazards (t(28) = 2.0, $p = 0.05$), though they clearly out-perform the novices on EP hazards (t(28) = 4.5, $p < 0.001$).

A second 2 x 2 mixed Analysis of Variance (ANOVA) was conducted on experienced and novice drivers' confidence ratings for BP and EP clips. There was a main effect of experience (F(1,28) = 11.3, MSe = 0.4, $p < 0.005$) with experienced drivers reporting greater confidence that their predicted hazard was imminent compared to novice drivers (with mean ratings of 4.3 and 3.7, respectively, from a 5 point scale). There was also an effect of hazard type (F(1,28) = 10.9, MSe = 0.1, $p < 0.005$) with BP clips eliciting greater confidence than EP clips (4.2 vs. 3.9). The interaction was not significant. When non-hazard ratings were included in the ANOVA (creating a 2 x 3 analysis, with three levels of hazard type) it did not change the overall pattern of results, but it was noticeable in the simple contrasts on the main effect of hazard type, that confidence ratings to EP clips were actually lower than those for the non-hazard clips (F(1,28) = 4.9, MSe = 0.2, $p < 0.05$).

Discussion

These results again discriminate between novice and experienced drivers. The overall effect size is considerable (Cohen's d = 1.52), but the effect is modified by the interaction with hazard type. While BP hazards offer marginal evidence for a difference between the groups, the EP hazards provide a clear distinction between the two. In addition to supporting the hypothesis that hazard prediction is a key element of hazard perception, the results support the categorization of hazards according to the relationship between precursor and hazard, suggested by Crundall et al. (2012), and similar to the distinctions noted in Vlakveld et al., (2013), and Borowsky and Oron-Gilad (2013). It was previously argued that EP hazards should be harder to predict because of the indirect relationship between precursor and hazard. Whereas the precursor of a person on the pavement might prime the driver directly for the possibility that the pedestrian might step into the road, the occluding property of a parked vehicle is more tenuously linked to the possibility of a pedestrian entering the road. Thus an environmental prediction may require a conscious linking of the environment to explicit knowledge imparted by previous experience of similar scenarios in the real world. Unfortunately the eye movement data from Crundall et al. (2012) were not clear cut, suggesting that all groups of drivers fixated EP precursors relatively late (though experienced drivers and driving instructors were still faster and more likely to fixate the EP hazard than novice drivers). Either this means that the EP precursors are irrelevant to the detection of EP hazards, or that the more experienced drivers used peripheral vision to monitor EP precursors until proximity was such that they warranted a fixation.

The current results have demonstrated that the EP precursors *are* used by experienced drivers in order to maintain their predictive accuracy. Novice drivers' performance suffers greatly however when faced with EP precursors. This is possibly due to the less direct link between precursor and hazard in EP events compared to BP events, though it is also possible that novices have less experience with EP events. This might be because EP events are statistically rarer than BP events, or that current commercial HP training and testing materials in the UK focus primarily on BP events.

Exactly how experienced drivers made use of the current EP precursors cannot be answered from the current data. Crundall et al. (2012) found experienced drivers are less likely to fixate EP precursors than BP precursors, arguing for a role for peripheral monitoring. In contrast to this suggestion, Vlakveld et al. (2011) have found initial success in the creation of a simulator training process that emphasizes anticipatory fixations on covert hazard locations in young novice drivers (i.e. environmental precursors, such as the leading edge of a parked bus). However their assessment of success is based on whether training drivers to look at EP precursors does indeed result in such fixations, and not whether these fixations result in safer driving. Thus it remains unclear whether they were training the most appropriate visual strategy. A similar eye-training study undertaken in Germany using computer generated hazard clips has also demonstrated success in evoking anticipatory fixations on hazardous precursors, but also failed to assess the impact whether this change in visual search improved drivers’ responses to hazards (Petzoldt et al., 2013).

Without eye tracking participants while viewing the clips used in the current study it is impossible to tell whether the current experienced drivers fixated the EP precursors early, or merely monitored them via peripheral vision, but it provides a future avenue of research for our understanding of the hazard perception process.

Ratings of confidence suggested that all drivers felt less sure of their predictions for EP hazards, though in contrast to experiment 2 there was also a group effect with experienced drivers reporting greater confidence in their answers than novices.

General Discussion

The primary aim of this paper was to assess whether the predictive element of hazard perception could consistently discriminate between experienced and novice drivers across a series of experiments. It has been noted that traditional measures of hazard perception (primarily response time measures) have produced mixed results with some notable studies failing to identify driver group differences (e.g. Chapman and Underwood., 1998; Crundall et al., 1999, 2002; Sagberg and Bjornskau, 2006; Borowsky et al., 2010, Underwood, et al., 2013). While these failures to replicate may relate to differences in the stimuli used by various research groups, one other possibility is that the interplay of the various sub-components of a typical HP response may mask underlying differences. Response times to hazards are likely to include processing time, confirmation, appraisal and bias, in addition to the prediction of the imminent hazard. The current paradigm removes all these additional influences and focuses solely on the predictive element of hazard perception. All three experiments have successfully discriminated between novice and experienced drivers using this measure of prediction, with considerable effect sizes.

In addition to the success of overall group discrimination, the first experiment provided a potential insight into the supposed effortful nature of hazard perception, suggesting that the secondary task impairment noted by McKenna and Farrand (1999) might have acted upon post-prediction processes. While one cannot conclude from this study that hazard perception is effortless, it certainly appears that experienced drivers can perform at a higher level for a longer period than novices, suggesting that there is a resource-based benefit of experience for at least the predictive component of hazard perception.

The second and third experiments investigated the nature of the precursors. Experiment 2 demonstrated that, though temporal distance between precursor and hazard reduces the cues available to prediction, experienced drivers were still able to maintain superior performance. This suggests that experienced drivers can extract relevant information from precursors more than a second before the hazardous event occurs. Thus the early fixations on some precursors noted by Crundall et al., (2012) may indeed have contributed to the hazard perception process, rather than just reflecting a general search strategy.

Finally, Experiment 3 compared prediction across two types of hazards: environmental prediction hazards and behavioural prediction hazards. The indirect link between precursor and hazard was possibly the reason that participants performed more poorly on EP hazards, though these clips were better at discriminating between novice and experienced drivers. This supports other research that has demonstrated that experienced drivers pay more attention to environmental context when evaluating hazardousness than novices (Borowsky and Oron-Gilad, 2013).

In addition to informing the theoretical underpinning of hazard perception, the results have also raised pragmatic issues regarding hazard perception testing as part of an official licensing procedure. For instance the length of the clips used in typical hazard perception tests has not been considered in the research literature. The results of experiment 1 suggest that clip length, possibly through the depletion of resources, can have a significant impact on inexperienced drivers' performance in predicting hazards (which would presumably also have an impact in a traditional HP test). Also the nature of the precursor-hazard relationship appears crucial to the probability of successfully predicting a hazard. While the results do not necessarily argue for an HP test based solely on EP events (for we have not tested the discriminative validity of EP events in isolation), they at least argue for a balanced mix of event types.

Perhaps the most impressive pragmatic message is how robust the discrimination effect sizes are when based solely on hazard prediction. Does this mean that hazard perception testing should abandon the current traditional method in favour of *hazard prediction*? There are three caveats which prevent one drawing conclusion immediately. First, these effects need to be replicated by other researchers. The niggling doubt surrounding traditional HP tests is, in part at least, derived from those studies that have failed to replicate the basic effect. Hopefully researchers will take up the challenge to investigate whether hazard prediction is a more robust discriminator than traditional HP methods.

Secondly, the test as it stands is not suitable for testing thousands of drivers as part of a licensing procedure as it requires participants to give relatively detailed verbal or written answers. Greater automation is possible however, and emerging results suggest that providing drivers with multiple-choice options at the end of the trial can preserve the discriminant validity (Castro et al., 2014).

Finally, it should be noted that this research is primarily driven by curiosity of what sub-processes are engaged in a hazard perception task. It has focused on prediction as the likely first step in a long chain of processes that eventually results in a typical HP RT response. It does not follow however that a measure of hazard prediction should necessarily be used in isolation to assess driver competence. Post-prediction processes may be equally important in separating the safe drivers from the unsafe ones, though until we have developed a more thorough understanding of the HP process we will not be able to identify which sub-components are the most important.

References

Baddeley, A. D., Hitch, G. J., and Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language, 61*, 438-456.

Bolstad, C. A., Cuevas, H. D., Wang-Costello, J., and Endsley, M. R. (2010).Measurement of situation awareness of automobile technologies of the future. In G. L. Rupp (Ed.) *Performance metrics for assessing driver distraction.* SAE International

Borowsky, A., Shinar, D., and Oron-Gilad, T., (2010). Age, skill, and hazard perception in driving. *Accident Analysis and Prevention, 42, 4,* 1240-1249.

Borowsky, A., and Oron-Gilad, T. (2013). Exploring the effects of driving experience on hazard awareness and risk perception via real-time hazard identification, hazard classification, and rating tasks. *Accident Analysis and Prevention*, *59*, 548-565.

Castro, C., Padilla, J. L., Roca, J., Benítez, I, García-Fernández, P., Estévez, B., López-Ramón, M.F., and Crundall, D.(2014). A first development and validation of the Spanish Hazard Perception Test. *Traffic Injury Prev*ention, DOI:10.1080/15389588.2013.879125.

Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: danger and experience. *Perception, 27*, 951-964.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.),Visual information processing (pp. 215–281). New York: Academic Press.

Chi, M.T.H., Feltovich, P.J., Glaser, R., 1981. Categorization and representation of physics problems by experts and novices. *Cognitive Science 5,* 121-152.

Cohen, J. (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Crundall, D. Bibby, P., Clarke, D. D., Ward, P., & Bartle, C. (2008). Car drivers' attitudes towards motorcyclists: a survey. *Accident Analysis and Prevention, 40, 3,* 983-993.

Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others. *Accident Analysis and Prevention, 45,* 600-609.

Crundall, D., Underwood, G., and Chapman, P. (2002). Attending to the peripheral world while driving. *Applied Cognitive Psychology, 16,* 459-475.

Crundall, D. and Underwood, G. (1999). Is attention required in a model of saccade generation? *Behaviour and Brain Sciences, 22,* 679-680.

Divekar, G., Pradhan, A. K., Pollatsek, A., and Fisher, D. L. (2012). Effect of external distraction behavior and vehicle control of novice and experienced drivers evaluated. *Transportation Research Record, 2321,* 15-22.

Deery, H. A., (1999). Hazard and risk perception among young novice drivers. *Journal of Safety Research, 30, 4,* 225-236.

Durso, F. T., Bleckely, M. K., and Dattel, A. R. (2006). Does situational awareness add to the validity of cognitive tests? *Human Factors, 48, 4,* 721-733.

Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 137–154). Hampshire, UK: Ashgate.

Durso, F. T., & Gronlund, S. D. (1999). Situation awareness. In T. F. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *Handbook of applied cognition* (pp. 283–313). New York: Wiley.

Durso, F. T., Truitt, T. R., Hackworth, C., Crutchfield, J., Nikolic, D., Moertl, P., et al. (1995). Expertise and chess: A pilot study comparing situation awareness methodologies. In D. J. Garland & M. R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 295–304). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Endsley, M. R., (1988a). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society 32nd Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M.R., (1988b). Situation awareness global assessment technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference* (pp789–795). New York: IEEE.

Endsley, M.R., (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors 37*, 32–64.

Ericsson, K. A., & Kintsch, W., (1995). Long-term working memory. *Psychological Review, 102,* 211–245

Gobet, F., & Simon, H. A., (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31,* 1–40.

Grade, S., Lefèvre, N., and Pesenti, M., (2013). Influence of gaze observation on random number generation. *Experimental Psychology, 60, 2,* 122-30.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., Parasuraman, R. (2003). Vigilance decrement reflects limitations in effortful attention vigilance decrement reflects limitations in effortful attention. *Human Factors, 45, 3*, 349-359.

Groeger, J.A. (2000). *Understanding Driving; Applying cognitive psychology to a complex everyday task.* Psychology Press, Routledge.

Gutzwiller, R. S., and Clegg, B. A. (2013). The Role of Working Memory in Levels of Situation Awareness. *Journal of Cognitive Engineer and Decision Making, 7, 2,* 141-154.

Gugerty, L. (2011). Situation awareness in driving. In D. L. Fisher, M. Rizzo, J. K. Caird and J. D. Lee (Eds.), *Driving simulation for engineering, medicine, and psychology* (pp 19.1–19.9). Boca Raton, FL: CRC Press.

Helton, W. S., and Russell, P. N. (2011). Working memory load and the vigilance decrement. *Experimental Brain Research, 212, 3,* 429-437.

Horswill, M.S., and McKenna, F.P., (2004). Drivers' hazard perception ability: situation awareness on the road. In: Banbury, S., Tremblay, S. (Eds.), *A Cognitive Approach to Situation Awareness* (pp. 155–175). Aldershot, UK: Ashgate.

Jackson, A. L., Chapman, P., & Crundall, D. (2009). What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics, 52, 2*, 154-164.

Jeannot, E., Kelly, C., Thompson, D. (2003). *The development of situation awareness measures in ATM systems.* EATMP report. HRS/HSP-005-REP-01.

Johannsdottir, K. R., and Herdman, C. M. (2010). The role of working memory in supporting drivers' situation awareness for surrounding traffic. *Human Factors, 52, 6,* 663-673.

Kim, H.S., Choi, M. H., Choi, J. S., Kim, H. J., Hong, S. P., Jun, J. H., Tak, G. R., Kim, B., Min, B. C., Lim, D. W. and Chung, S. C. (2013). Driving performance changes of middle-aged experienced taxi drivers due to straction tasks during unexpected situations. *Perceptual and Motor Skills, 117, 2,* 411-426.

Lim, P. C., Sheppard, E., and Crundall, D. (2014). A predictive hazard perception paradim differentiates driving experience cross-culturally. *Transportation Research Part F, 26,* 210-217.

McKenna, F. P., & Crick, J. L. (1991). *'Hazard Perception in Drivers: a Methodology for Tetsing and Training', Final Report*. Transport Research Laboratory. Crowthorne, UK.

McKenna, F., & Crick, J. (1997). Developments in hazard perception. TRL report (TRL297) Crowthorne, UK: Transport Research Laboratory.

McKenna, F.P. & Farrand, P. (1999). The role of automaticity in driving. In GB Grayson (Ed.) Behavioural Research in Road Safety, 9, Transport and Road Research Laboratory, Crowthorne.

McKenna, F. P. & Horswill, M. S. (1999). Hazard perception and its relevance for driver licensing. Journal of the International Association of Traffic and Safety Sciences, 23,1, 36-41

Pelz, D. C., & Krupat, E. (1974). Caution Profile and Driving Record of Undergraduate Males. *Accident Analysis and Prevention, 6*, 45-58.

Petzoldt, T., Weiß, T., Franke, T., Krems, J.F., & Bannert, M. (2013). Can driver education be improved by computer based training of cognitive skills? *Accident Analysis and Prevention, 50*, 1185-1192.

Pradhan, A. K., Hammel, K. R., DeRamus, R., Pollatsek, A., Noyce, D. A., & Fisher, D. L. (2005). Using eye movements to evaluate effects of driver age on risk perception in a driving simulator. *Human Factors, 47*(4), 840-852.

Sagberg, F., & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. *Accident Analysis and Prevention, 38, 2*, 407-414.

Salmon., P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., and Young, M. (2009). Measuring situation awareness in complex systems: comparison of measures study. *International Journal of Industrial Ergonomics, 39*, 490-500.

Salmon., P. M., Young, K. L., and Cornelissen, M. (2013). Compatible cognition amongst road users: The compatability of driver, motorcyclist and cyclist situation awareness. *Safety Science, 56,* 6-17.

Savage, S. W., Potter, D. D., and Tatler, B. W., (2013). Does preoccupation impair hazard perception? A simultaneous EEG and eye tracking study. *Transportation Research Part F: Traffice Psychology and Behaviour, 17,* 52-62.

Sohn, Y. W., and Doane, S. M. (2004). Memory Processes of Flight Situation Awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. *Human Factors 46, 3,* 461-475.

Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D., (2005). *Human Factors Methods: A Practical Guide for Engineering and Design.* Ashgate, Aldershot.

SWOV, (2010). Hazard perception in traffic. SWOV fact sheet. Leidschendam, The Netherlands: SWOV.

Taylor, R. M., (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. In Situational Awareness in Aerospace Operations (AGARDCP-478) pp3/1 –3/17. Neuilly Sur Seine, France: NATO-AGARD.

Underwood, G., Chapman, P., Bowden, K. & Crundall, D. (2002). Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F, 5*, 87-97.

Underwood, G., Ngai, A., and Underwood, J. (2013). Driving experience and situation awareness in hazard detection. *Safety Science, 56,* 29-35.

Vlakveld, W., Romoser, M., Mehranian, H., Diete, F., Pollatsek, A., and Fisher, D. (2011). Do crashes and near crashes in simulator-based training enhance novice drivers' visual search for latent hazards? 2265, 153-158.

Vidulich, M., Dominguez C., , Vogel, E. and McMillan, G. (1994). Situation awareness: papers and annotated bibliography. Technical Report No. AL/CF-TR-1994-0085. Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Vogel, K., Kircher, A., Alm, H., & Nilsson, L. (2003). Traffic sense-which factors influence the skill to predict the development of traffic scenes? *Accident Analysis and Prevention, 35*(5), 749-762.

Walker, G. H., Stanton, N. A., Kazi, T. A., Salmon, P. M., and Jenkins, D. P. (2009). Does advanced driver training improve situational awareness? *Applied Ergonomics, 40,* 678-687.

Warm, J. S., Parasuraman, R., and Matthews, G. (2008). Vigilance requires hard mental work and is stressful. Human Factors, 50, 3, 433-441.

Watts, G. R., & Quimby, A. R. (1979). *'Design and validation of a driving simulator for use in perceptual studies', Report 907*. Transport Research Laboratory. Crowthorne, UK.

Wetton, M. A., Hill, A., Horswill, M. S. (2013). Are what happens next exercises and self-generated commentaries useful additions to hazard perception training for novice drivers? Accident Analysis and Prevention, 54,57-66.

Wetton, M. A., Horswill, M., Hatherly, C, Wood, J. M., Pachana, N. A., and Anstey, K. J., (2010) The development and validation of two complementary measures of drivers' hazard perception ability. Accident Analysis and Prevention, 42, 4, 1232-1239.

Wetton, M. A., Hill, A., and Horswill, M. S. (2011). The development and validation of a hazard perception test for driver licensing. *Accident Analysis and Prevention, 43, 5,* 1759-1770.

Figure Legends:

Figure 1. Three frames depicting the final few seconds before the image is occluded by a black screen. The pedestrian becomes visible in the top panel, moves towards the crossing in the middle panel, and is about to step onto the crossing in the third panel (immediately prior to occlusion). These frames were also the different end points for three versions of this clip used in Experiment 2. The top panel represents the earliest end point (i.e. shortest clip length), where a pedestrian’s head can first be seen over the top of a parked vehicle on the left. The middle panel shows the pedestrian moving towards the zebra crossing; this is the intermediate end point. The bottom panel shows the late end point. The pedestrian is now completely visible and is about to step into the road.

Figure 2. Participants’ mean accuracy in predicting the upcoming hazard according to clip length (with standard error bars added).

Figure 3. Top panel: Participants’ mean accuracy scores for predicting the upcoming hazard across the three clip end points (with standard error bars added). Bottom panel: Participants’ mean confidence ratings for their responses to the upcoming hazard (on a scale from 1-5, where 1 is extremely low confidence and 5 is extremely high confidence that the hazard will occur; with standard error bars added). The dashed line represents participants’ average confidence in their answers to non-hazardous clips.

Figure 4. A Behavioural Prediction video clip. Towards the end of the clip participants are ‘driving’ on a three-lane road through Nottingham City. The white van moves into the left lane (top panel) while a parked car pulls off from the right (middle panel). The final frame of

the clip prior to the screen turning black (bottom panel) shows the white van indicating to return to the middle lane (thus causing a hazard). This can be predicted on the basis of the movement of the white van prior to indicating, as it drifts towards the centre lane, and the fact that parked vehicles in the white van’s lane are visible suggesting that he will have to move out (circled in top panel).

Figure 5. Participants’ percentage accuracy scores for predicting subsequent hazards (with standard error bars added).

Figure 1.

Figure 2.

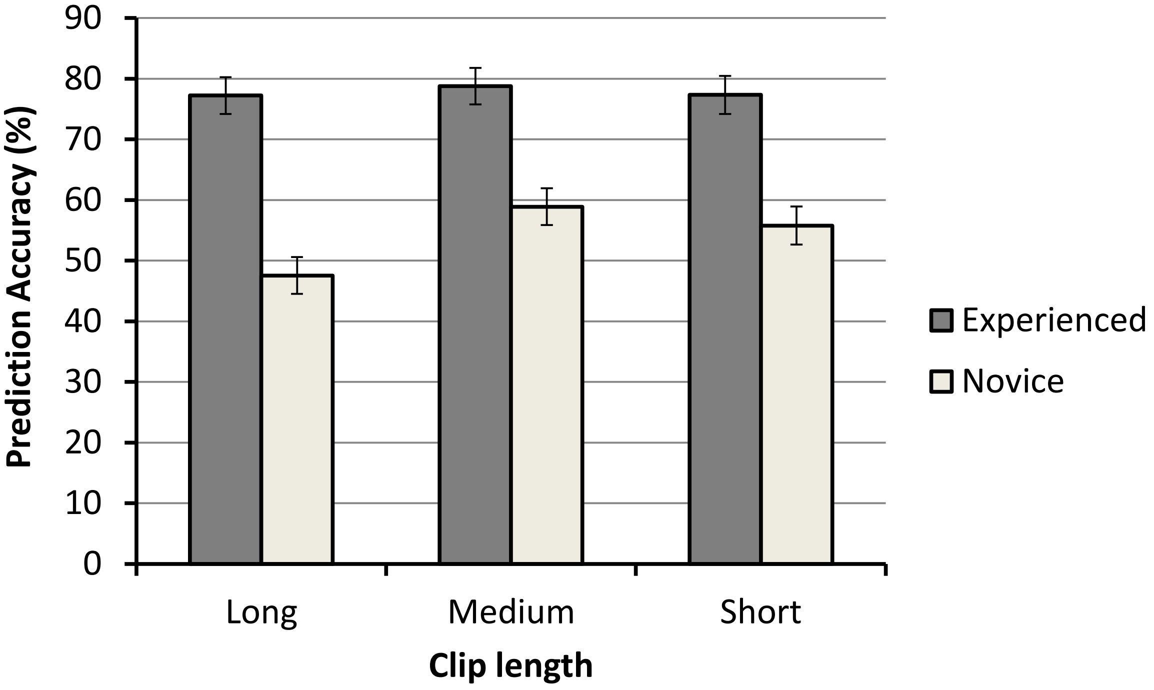
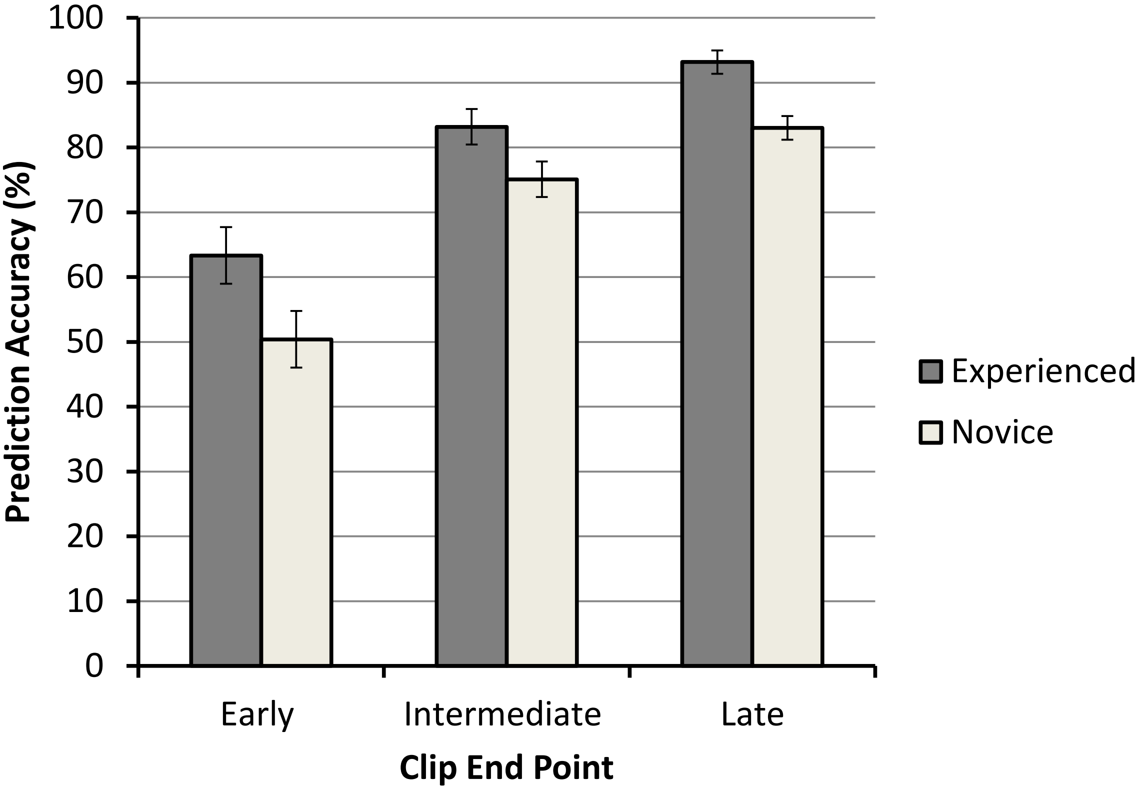

90
80
70
60
50
40
30
20
10
0
Prediction Accuracy (%)
Long
Medium
Short
Clip length
Experienced
Novice

Figure 3.


100
90
80
70
60
50
40
30
20
10
0
Prediction Accuracy (%)
Early
Intermediate
Late
Clip End Point
Experienced
Novice

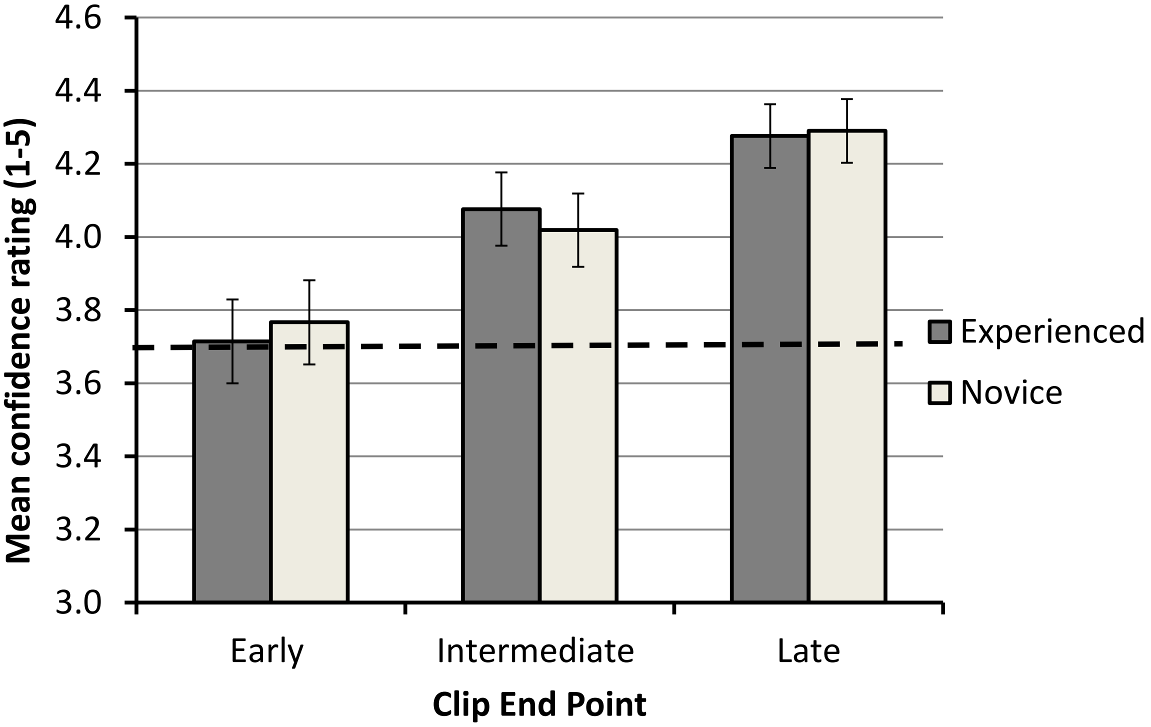
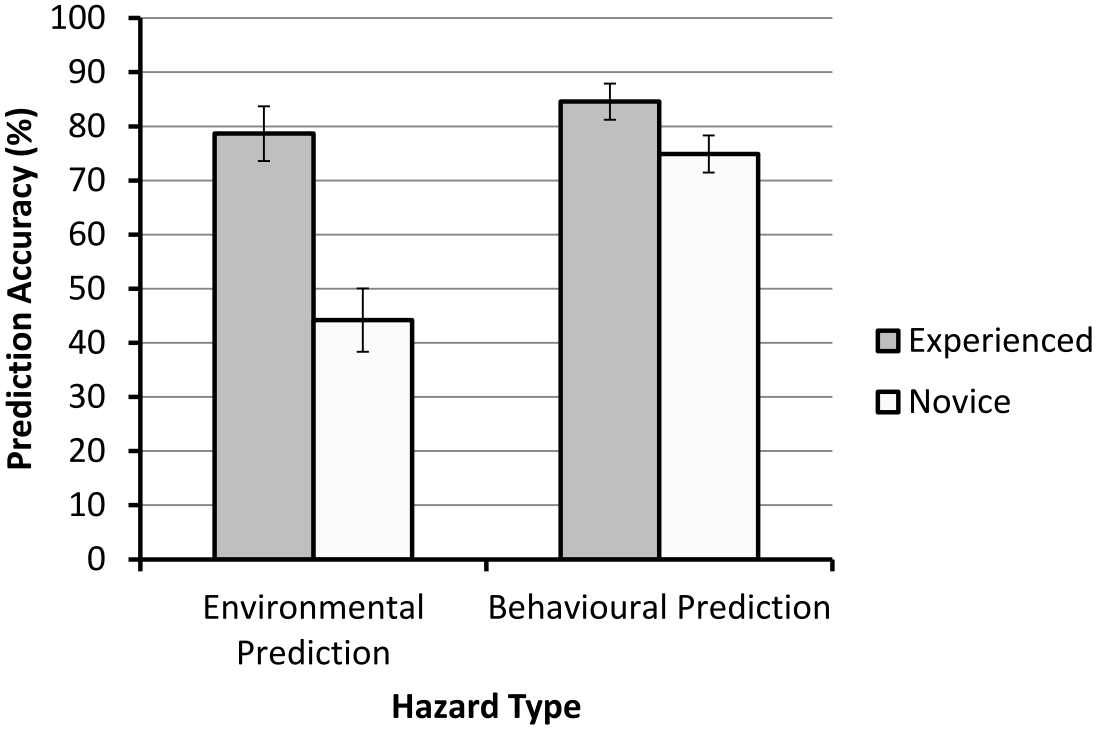

4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
Mean confidence rating (1-5)
Early
Intermediate
Late
Clip End Point
Experienced
Novice

Figure 4.

Figure 5


100
90
80
70
60
50
40
30
20
10
0
Prediction Accuracy (%)
Environmental Prediction
Behavioural Prediction
Hazard Type
Experienced
Novice